4-2015

Exploring cyberbullying and other toxic behavior in team competition online games

Haewoon KWAK
Singapore Management University, hkwak@smu.edu.sg

Jeremy BLACKBURN

Seungyeop. HAN

Citation

1

Exploring Cyberbullying and Other Toxic Behavior in Team Competition Online Games

Haewoon Kwak
QCRI
Doha, Qatar
hkwak@qf.org.qa

Jeremy Blackburn
Telefonica Research
Barcelona, Spain
jeremyb@tid.es

Seungyeop Han
University of Washington
Seattle, WA, USA
syhan@cs.washington.edu

ABSTRACT

In this work we explore cyberbullying and other toxic behavior in team competition online games. Using a dataset of over 10 million player reports on 1.46 million toxic players along with corresponding crowdsourced decisions, we test several hypotheses drawn from theories explaining toxic behavior. Besides providing large-scale, empirical based understanding of toxic behavior, our work can be used as a basis for building systems to detect, prevent, and counter-act toxic behavior.

Author Keywords

MOBA; online video game; toxic playing; team competition; cyberbullying; crowdsourcing; League of Legends; trolling

ACM Classification Keywords

J.4 Computer Applications: Social and Behavioral Sciences—*Sociology, Psychology*; K.4.2 Computers and Society: Social Issues—*Abuse and crime involving computers*

INTRODUCTION

With the remarkable advances from isolated console games to massively multi-player online role-playing games (MMORPG), the online gaming world provides yet another place where people interact with each other. The main reasons that researchers pay attention to online games are 1) that the purpose of actions is relatively clear, and 2) that actions are quantifiable. A wide range of predefined actions for supporting social interaction (e.g., friendship, communication, trade, enmity, aggression, and punishment) reflects either positive or negative connotations among game players [47], and they are unobtrusively recorded by game servers. These rich electronic footprints enable and encourage research of social dynamics [22, 28, 47].

There is no end to the growth of online gaming in sight. For example, in 2011 the First Dota 2 International Tournament had a prize pool of $1,600,000. In 2014, the prize pool started again at $1.6M, but, a *crowdfunding* mechanism brought it up to just under $11M. For comparison, the winners of the 2014 International received $5M ($1M for each team member), while the winner of the 2014 US Open Men's tennis

tournament received $3M. In other words, video games have reached a level of sophistication and competitiveness that not only can gamers make a living playing them, they can make quite a comfortable living.

While the widely adopted game design element of competition increases enjoyment [52], it has also led to an increasing concern about negative behavior. In online gaming, negative behavior, such as cyberbullying [46], griefing [26], mischief [31], and cheating [10] are often grouped together and called *toxic behavior*. Unfortunately, the definition of toxic behavior is often unclear [16] due to differences in expected behavior, customs, rules, and ethics across games [53]. Such ambiguity and subjective perception of griefing make griefers themselves sometimes fail to recognize what they did [34].

To make matters worse, since online games are very popular in the younger generation [5], instances of cyberbullying can cause far reaching problems. In general, cyberbullying is associated with depression, anxiety, and has been shown to result in drastic actions such as suicide in several well publicized cases [6]. With the amount of time and energy players invest into games, victims of toxic behavior are likely to feel emotional effects that persist to the real-world.

In this paper, we make use of a large-scale dataset capturing millions of instances of toxic behavior perpetuated by hundreds of thousands of accused toxic players from the League of Legends (LoL)[1], the world's most popular online game [4]. The richly detailed cases are augmented by crowdsourced decisions on whether or not the accused was in fact toxic. Drawing from sociology and psychology literature, we explore toxic behavior through the lens of competitive online games.

BACKGROUND

In this section we begin with a basic description of LoL and the Tribunal, its crowdsourcing platform for addressing toxic behavior. We highlight the differences between LoL and other online games from the perspective of team formation, the goal of the game, and the common game modes. We then move on to how Riot Games attempts to detect toxic players and which behavior is considered toxic. Lastly, we explain how the Tribunal system works.

League of Legends as a Team Competition Game

The League of Legends (LoL), a Multiplayer Online Battle Arena (MOBA), is arguably the most popular online game in the world. The developer, Riot Games, recently announced

[1]http://www.leagueoflegends.com

there are 67 million players per month, 27 million players per day, and over 7.5 million concurrent players at peak times [2]. LoL is an online team competition game whose goal is to penetrate and destroy the enemy's central base, called the Nexus. In contrast to massively multiplayer online role-playing game (e.g., World of Warcraft), LoL is a match-based team competition game, where a single match usually takes around 30 to 40 minutes. Although LoL provides a few different modes of matches, all the modes are competitions between two teams. In the most popular game mode, each team has five members. A player is given the option to form a team with friends before the match. Otherwise, the player is randomly assigned to a team with strangers who have similar skill levels.

Player Reports on Toxic Behavior

After a match, players see a scoreboard as in Figure 1. In Figure 1, (A) is the match summary, (B) lists players who played the game together, and (C) is a chat window. Users can report toxic players by clicking the rightmost red button in (B). Each player can submit one report for every other player per game. We stress that player reports can only be submitted *after* the match is completed, and at a later date reviewers vote as described in the next section. Therefore, player reports are not considered a strategic element during the match itself.

Figure 1. LoL scoreboard after a match

When reporting, players choose from several predefined categories of toxic playing: assisting enemy team, intentional feeding [suicide], offensive language, verbal abuse, negative attitude, inappropriate name, spamming, unskilled player, refusing to communicate with team, and leaving the game/AFK [away from keyboard]. Detailed explanations of each type of toxic playing are given in [3]. Each report is submitted under one category. We further classify these categories as either cyberbullying (offensive language and verbal abuse) or domain specific toxicity (the remaining categories).

While most of categories are self explanatory, two of the domain specific categories need additional explanation: intentional feeding and leaving the game. Intentional feeding indicates that the player died to the opposing team on purpose, often repeatedly. Leaving the game/AFK refers to instances when a user is inactive through the entire match. These two categories of toxic behavior usually make the opposing team stronger due to the game's design.

To understand the motivation behind intentional feeding and leaving the game, we present one common scenario where such toxic play happens. A LoL match is concluded when the Nexus of one team is destroyed. If one team is obviously at a disadvantage during the game, the team may give up the match by voting to "surrender". However, surrender is accepted only when at least four out of five players on the same team agree to do so. When the surrender vote fails, players who voted for surrender may lose interest in continuing the game. Then, they might exhibit extreme behavior, such as intentional feeding or leaving the game/AFK, in an attempt to force the game to finish earlier or convince other players to cast a vote for surrender.

LoL Tribunal as a Crowdsourcing System

The LoL Tribunal is a crowdsourcing system to make decisions on whether reported players should be punished or not. Only players that are reported more than a few hundred times are brought to the Tribunal [11]. Once this threshold is reached, up to 5 randomly selected matches that an accused player was reported in are aggregated into a case. Reviewers, who are LoL players with enough time invested to be considered "experts," are then presented with a rich set of information about the case. They see the full chat logs from each match, the performance (similar to the end score board) of each player in the game, the duration of the match, and the category (and optional comments) that the player was reported for. To ensure an unbiased result, all players are anonymized in the match data. We note that among 10 predefined categories of toxic behavior, the Tribunal excludes reports of unskilled player, refusing to communicate with team, and leaving the game.

It is known that about 100-150 reviewers cast votes for a single case [11], and a majority voting scheme is used to reach a verdict. Guilty verdicts can result in a ban for 1 week, a few months, and sometimes longer suspensions/permanent bans. E.g., professional players have received lengthy, career impacting suspensions due to Tribunal decisions [1].

DATA COLLECTION

We collect all information available in the Tribunal: player reports, the crowdsourced decision, and detailed match logs. This rich collection is the basis for our quantitative analysis.

Early studies on moral behavior in sports largely depend on self-reports [41], but they suffer in quality and scalability from bias in the human recall process [35]. Furthermore, self-reports in particular can be influenced by social desirability, the tendency to behave in a way that is socially preferable [29]. In this sense, reports submitted by 3rd parties, as in the Tribunal, are more reliable since they eliminate the bias of self-report.

Another widely used method is asking individuals about hypothetical situations, but it also has the same social desirability bias, and additionally has limitations due to the specificity of scenarios and scalability [42]. Our huge collection of witness reports on *observed* toxic behavior helps address these problems. Furthermore, crowdsourced decisions offer an even more objective viewpoint on reported toxic behavior.

Riot Games divides the world into several regions and maintains dedicated servers for each region. We focus on three highly active regions, North America (NA), Western Europe (EUW), and South Korea (KR) by considering representativeness of cultural uniqueness and familiarity to the authors. Although a player may connect to servers operated in different regions, a further distance between a player and a server usually results in increased latency in Internet connections, sacrificing the quality of responsiveness and interactiveness of online games. We thus reasonably assume most players connect to the servers corresponding to their real-world region for the best quality of service.

In April 2013, we collected about 11 million player reports from 6 million matches on 1.5 million potentially toxic players across three regions. We collect all available data from the servers and summarize it in Table 1. We first note that the KR portion of our dataset is smaller than other regions because the KR Tribunal started in November 2012 but the EUW and NA Tribunals started in May 2011. Next, since player reports are internally managed, it is not easy to measure our dataset's completeness. However, as reviewers can see their past votes and final decisions at a later time, it makes sense that no votes or reports are removed from servers and we have what amounts to a complete picture of the Tribunal.

	NA	EUW	KR	TOTAL
Player	590,311	649,419	220,614	1,460,344
Match	2,107,522	2,841,906	1,066,618	6,016,046
Report	3,441,557	5,559,968	1,893,433	10,898,958

Table 1. Summary of our dataset.

RESEARCH QUESTIONS AND HYPOTHESES

In this section we formulate our research questions on how a toxic player behaves as well as how other players react to the toxic player.

At its core, LoL is a team competition game. Although competition against human opponents provides a high degree of player enjoyment [54], it can also result in toxic behavior. The potential for toxic behavior can be further exacerbated by the anonymous aspects of LoL, as people do not feel accountable for their toxic behavior when anonymous and can actually increase their aggressive behavior [17]. We present several compelling theories from sociology and psychology that describe the "whys" of toxic behavior. Our large-scale dataset gives the opportunity to apply these theories "in the wild."

We begin with an explanation of the bystander effect and the vague nature of toxic playing. Next, we discuss the concepts of in-group favoritism and out-group hostility with respect to competition. We then describe how intra-group conflict emerges in a team competition environment like LoL and some related theories which indicate a socio-political effect on how toxic behavior is both expressed and perceived. Finally, we consider the effects of team-cohesion on performance, which can provide insights into what might trigger toxicity in online games in particular.

Bystander Effect and Vague Nature of Toxic Playing

Our first research question is focused on how actively people report toxic playing. It is essential to understand the effectiveness and design implications of a system for counter-acting toxic playing, such as the Tribunal.

RQ1a: How active are players in reporting toxic behavior?

This research question is heavily related to the social psychology concept of the bystander effect [9], which describes the tendency for observers to avoid helping a victim, particularly when they are immersed in a group. Considering LoL's anonymous setting with ephemeral teams, the bystander effect can influence how other players react to a toxic player. If the bystander effect is valid in this setting, then most players in a match will *not* report the toxic player even though they directly witnessed the abuse. We explore how actively players report their observation of toxic behavior via our dataset.

Interestingly, the bystander effect is known to be mitigated with explicit pleas for aid [33]. We can thus draw a testable hypothesis.

HYPOTHESIS 1.1. *If there is a request asking to report toxic players, the number of reports is increased.*

RQ1b: How does the vague nature of toxic playing affect tribunal decisions?

Additionally, players have different perceptions on the severity of toxic behavior and thus sometimes fail to recognize its presence [34]. This vague nature of toxic playing might affect both reporting and reviewing. We expect that different perceptions between players who report toxic behavior and Tribunal reviewers will result in a number of pardons.

In-group Favoritism and Out-group Hostility

The next question we explore is related to the team-based, competitive gameplay of LoL.

RQ2: What is the difference between reporting behavior of the toxic player's teammates and his opponents?

In-group favoritism is simply the tendency of people to favor in-group members (e.g., teammates) over similarly likable out-group members (e.g., opponents) while *disliking* out-group members when compared to similarly dislikable in-group members [51]. This is also related to homophily [37], which is the tendency for similar individuals to form relationships.

Deindividuation theory explains how an individual loses the concept of both self and responsibility when immersed in a crowd [21], which provides a possible mechanism behind in-group favoritism. This foundation meshes with the unique characteristics of computer-mediated communication (CMC): anonymity and reduced self-awareness [15, 48].

Reicher et al. [43] proposed the Social Identity Model of Deindividuation Effects (SIDE) as a way of explaining effects that classic deindividuation theory could not. In short, they discovered that simply being part of an anonymous crowd was insufficient impetus for displaying anti-normative behavior. Instead, in-group identity alone becomes relevant only

in comparison to a relevant out-group. In other words, both anonymity *and* context drive deindivduation.

Even though the groupings in LoL are ephemeral and mostly anonymous, the two teams are in direct competition and only one team can win: the thrill of victory or the agony of defeat is shared by all members of a team. This configuration catalyzes group identification and provides clear cut divisions of in- and out-groups in a context that is *driven* by out-group hostility; i.e., players are literally attempting to defeat the opposing team. Thus, we expect to see empirical evidence of in-group favoritism in reporting of some toxic behavior that equally affects both teams.

HYPOTHESIS 2.1. *For toxic behavior that affects both teams equally, in-group members (teammates) are less likely to submit reports when compared to out-group members (opponents).*

Intra-group Conflicts and Socio-political Factors

Although we expect to find evidence of in-group favoritism and out-group hostility, the team competition setting of LoL might also lead to substantial intra-group conflict. Intuitively, it is easy to blame overall team performance on a single poorly performing player. Unlike the aforementioned concept of in-group favoritism, the poorly performing player does not affect both teams equally because he makes the opponent team relatively stronger. Thus, he becomes a target to blame.

According to the classification scheme of human society introduced by Tönnies [49], teams in LoL are closer to task-oriented associations (Gesellschaft) than the social community associations (Gemeinschaft). In task-oriented associations, the relationship among players is somewhat impersonal and social bonding does not necessarily exist. Thus, toxic players might not feel a sense of a team and feel no qualms about harassing teammates who are hurdles to winning rather than recognizing enemies for beating his team.

While it is difficult to make direct conclusions about intra-group conflict, LoL players are generally segregated by what amounts to socio-political regions. The large world-wide user-base of LoL makes it feasible to study regional differences of such intra-group conflicts. Thus, leading to our next research question.

RQ3: What is the impact of socio-political factors on toxic behavior?

There are several studies to support that the degree to which bullying occurs is influenced by socio-political factors [8, 14]. Chee reports on a unique Korean gaming culture called *Wang-tta* [14]. In the context of gaming, *Wang-tta* describes the phenomenon of "isolating and bullying the worst game player in one's peer group." *Wang-tta* is thought to be modeled after the similar Japanese term *Ijime* [8] which describes the comfort members of collectivist societies feel from similarity and the abuse thrust upon those that are different.

Conversely, other socio-political regions such as North America and Western Europe tend to be individualistic, with a focus on relying on ones' self [27]. In such socio-political

regions, while bullying of poorly performing players certainly occurs, there is less ingrained hostility towards another player's poor individual performance [39]. For example, consider the counterpart to the cliché "there's no 'I' in team," "well there ain't no 'WE' either." There is simply more focus on *my* performance as opposed to *our* performance. If there are in fact socio-political factors at play, then we would expect to see this reflected in reports on toxic behavior from different regions, leading to a testable hypothesis.

HYPOTHESIS 3.1. *Due to a more group-success oriented socio-political environment, cyberbullying offenses are less likely to be punished in Korea than in other regions.*

Collectivist societies such as Korea and Japan have a desire towards similarity, while differences are a source of disparagement [39]. In addition, collectivism, by definition, places more emphasis on cooperation, the group goal, and a sense of belonging than individualism does [18, 50]. That is, deliberately harming the group is anathema to the highly held social value of cooperation and is met with intense derision. Thus, we propose the next two hypotheses from the perspective of reporters and reviewers.

HYPOTHESIS 3.2. *Reports on toxic behavior that largely affects the result of the match are more often submitted in Korea than in other regions.*

HYPOTHESIS 3.3. *Reports on behavior that largely affects the result of the match are more likely to be punished in Korea than in other regions.*

Team-cohesion and Performance

The cohesion-performance relationship in sports has been studied for decades. Several studies have confirmed a moderate but significant effect of cohesion on performance, with the effect varying according to the type of sport, the mechanism of team building, and the gender of players [13, 38]. More generally, Felps et al. discuss how a single negative member can bring group-level dysfunction [25]. The negative member violates interpersonal social norms and thus can lead to negative emotions and reduced trust among teammates. These psychological states then trigger defensive reactions and influence the overall functioning of the group. Intuitively, toxic behavior is likely to have a negative effect on team-cohesion, and thus performance, leading to our next research question.

RQ4: What is the relationship between toxic behavior, player reports, and team performance?

Weiner proposes attribution theory which states that individuals are likely to search for causal factors of failure, considering even innocuous factors as significant [55]. Naquin and Tynan present a similar concept, the team halo effect [40], which describes the tendency of people to give credit for success to the team as a collective, but to blame other people for poor team performance.

The underlying cognitive process is known as counterfactual thinking [44], which is when we create a mental simulation built on events that are contrary to the facts of what really happened. If an individual deduces a high likelihood of changing

the outcome to a more positive one via the mental simulation, then the difference between simulation and reality is identified as the causal factor for the negative outcome.

Generally speaking, counterfactual thinking helps us accurately identify causal factors of outcomes, but it can be biased by numerous factors [23], such as the halo effect. For example, if one player makes a poor decision, then a toxic player might imagine what would have happened if that decision was not made, attributing the current state of the match to that singular incident regardless of any mistakes he or other players might have made. Similarly, after the match is completed, a negative outcome and counterfactual thinking might lead bullied players to seek revenge on toxic players as a defensive reaction to perceived victimization and a method by which to gain some emotional satisfaction [20].

Since theory indicates negative outcomes trigger both toxic behavior and attempts to punish said toxic behavior, but reviewers of Tribunal cases are unbiased, we propose the following hypotheses.

HYPOTHESIS 4.1. *More reports come from matches where the accused was on the losing team.*

Also, such aggressive reporting might be likely to be pardoned.

HYPOTHESIS 4.2. *There are more cases pardoned when the accused was on the losing team than on the winning team.*

RESULTS

Bystander Effect and Vague Nature of Toxic Playing
First, we look into how actively people report toxic playing. When toxic play occurs in LoL, either 4 players (allies of the toxic player) or 9 players (all players but the toxic player) are exposed. For instance, if a toxic player is verbally abusive or uses offensive language in the ally chat mode, it is only visible to his allies. However, he can also use the all chat mode, exposing all 9 players to toxicity. For the other categories, all players are exposed in essentially the same manner.

Figure 2 plots the distribution of reports per match depending on the type of behavior reported for each region. Across all regions, the mean and median number of toxic reports per match are 1.812 and 1, respectively[2]. I.e., no more than 2 players per match report toxic playing on average. Overall, intentional feeding matches have the highest number of reports—about 50% of cases have more than 1 report, with over 60% for KR and EUW in particular—but the majority of matches have less than 3 reports. This is extremely low compared to the number of exposed players. I.e., on average, LoL players do not actively report toxic players.

To test *[H1.1] If there is a request asking to report toxic players, the number of reports is increased*, we look for situations where a request to report is sent to players on the enemy team; i.e., those that are not negatively affected by the toxic player, yet are likely to recognize his behavior and react if a plea is made. An explicit request sent to the opposite team when

[2]Every match in our dataset has at least one player report, as only players that have been reported appear in the Tribunal.
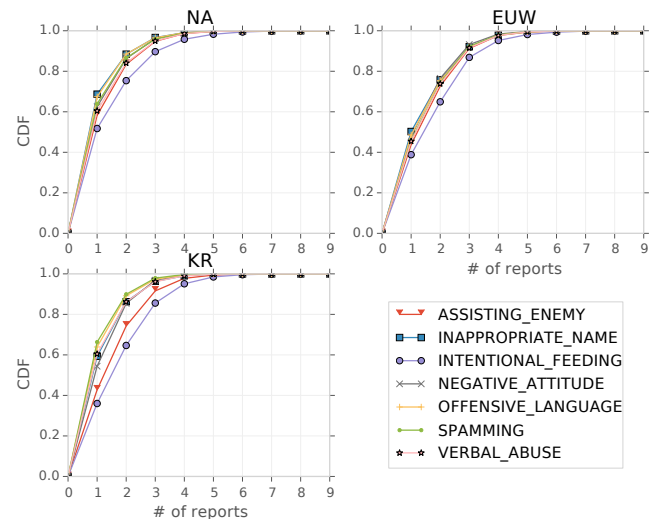

Figure 2. Distribution of the number of reports per match.

intentionally feeding or assisting the enemy satisfies this condition.

To test this, we define two dichotomous variables: 1) whether an explicit request exists, and 2) whether members of the opposite team report intentional feeding or assisting the enemy. We set the former variable to 1 when the word "report" appears in all chat, and 0 otherwise. We then create a 2×2 matrix of these variables. A Chi-Square test with Yates' continuity correction reveals that the percentage of reporting by enemies significantly differs with the existence of explicit requests ($\chi^2(1, N = 580,480) = 194552.9$, p < .0001).

Surprisingly, through the odds ratio we discover that the probability of opponents reporting is 16.37 times higher when allies request a report. This supports H1.1: explicit requests to report toxic players highly encourage enemies to report toxic players *even if toxic behavior is beneficial to the enemies*. In other words, we find the bystander effect is neutralized via explicit requests for help. This is interesting because opposing players can benefit from toxic playing, and their typical behavior of not reporting is changed due to explicit requests. This finding suggests that interaction design should actively encourage players to report others' toxic playing.

We next explore the possible impact of the vague nature of toxic playing on reporting and reviewing. First, we look into an association between the recognizability of toxic playing and the number of reports. Among the 7 types of toxic playing in LoL, intentional feeding and assisting the enemy team are, generally speaking, much more concrete expressions of toxicity than other types. The average number of reports per match for these two categories is higher than that for other types, as shown in Table 2. Both categories are consistently the first and the second ranked in terms of the average number of reports per match across all regions. A Kruskal Wallis test revealed a significant effect for category of toxic behavior on the number of reports per match ($\chi^2(6) = 117,399.1$, p < .0001). Further post-hoc statistical testing using the pairwise

Wilcoxon test with Bonferroni correction showed a significant difference between categories (p < .0001).

A.E	I.N	I.F	N.A	O.L	S	V.A
1.876	1.602	2.091	1.696	1.691	1.769	1.740

Table 2. Avg. number of reports per match for each type of toxic playing.
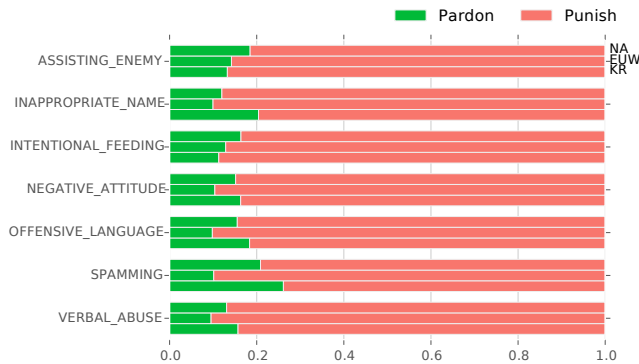

Figure 3. Proportion of decisions according to categories of toxic playing.

Second, we look at how many reported toxic players are pardoned in Figure 3. Since players come to the Tribunal only after a few hundred reports against them, we do not expect to see a 50/50 punish/pardon ratio. We obtained records for 477,383 punishments (80.9%) and 112,930 pardons (19.1%) for NA, 559,449 (86.0%) and 90,966 (14.0%) for EUW, and 187,253 (85.0%) and 32,929 (15.0%) for KR across all categories of toxic playing. The highest pardoned ratio we found for a specific category[3], spamming, is 26.1% in KR. Being reported means that *players* regard it as toxic playing, and a pardon means that *reviewers* do not regard it as toxic. Therefore, a 26% pardoned rate shows that different perceptions exist: for 1 in 4 cases, reviewers do not find the player toxic. This high pardoned ratio confirms the impact of the vague nature of toxic playing in reviewing as well as in reporting.

In-group Favoritism and Out-group Hostility
Our next research question is understanding the difference between reporting behavior of the toxic player's teammates and his opponents. To test *[H2.1] For toxic behavior that affects both teams equally, in-group members (teammates) are less likely to submit reports when compared to out-group members (opponents)*, we carefully revisit the definition of in-group favoritism. The key concept of in-group favoritism is "similarly" likable or unlikeable members from the same group are more favorable. We find this kind of relatively "neutral" toxic behavior from reports of inappropriate name. The inappropriate name of a toxic player is visible to all players equally, and thus the impact of the inappropriate name is neutral to both teams.

In Figure 4 we plot the number of matches reported for each category of toxic behavior per region. As we hypothesized,

---

[3]When multiple reports per accused player per game are submitted, we use the most common report type.
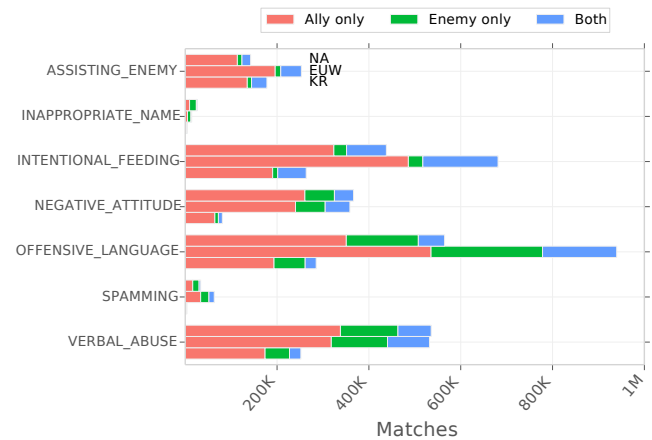

Figure 4. Number of matches that report each category of toxic behavior.

the only category in which the number of reports from enemies is higher than from allies is inappropriate name (ally only: 16,339, enemy only: 23,966, across regions). This indicates that allies are more forgiving, and thus less likely to report, than opponents in cases where the toxic player's offense is neutral to both teams. This is exactly what in-group favoritism describes, and thus H2.1 is supported.

Intra-group Conflicts and Socio-political Factors
We now explore our next research question on the relationship between intra-group conflicts and socio-political factors. Figure 4 shows the number of matches that are reported due to each category of toxic behavior. The horizontal bar for each category divided into three parts shows the number of matches reported by ally-only, enemy-only, and both, respectively from left to right. For instance, about 200 thousand matches are reported by ally-only due to assisting enemy, 12,106 matches by enemy-only, and 44,927 matches by both in EUW. We find that most reports come from allies rather from enemies as evidences of intra-group conflicts.

In Figure 4, offensive language is the most reported toxic behavior in LoL and is highly reported by allies. The chat feature is designed for exchanging strategies and sharing emotions, and thus serves to foster a sense of belonging in the team. In practice, however, it becomes a channel for toxic players to harass other ally players [32].

We additionally note that Riot Games disabled the all chat by default in newly installed clients since April 2012, while the Tribunal began in May 2011. Although players can easily turn the all chat option on in the game configuration by one click, we expect that it decreases the verbal interaction across teams after April 2012. To confirm that this is not the reason that players are likely to report allies rather than opponents for verbal abuse or offensive language, we look into the temporal trend of the toxic reports for those two reasons. We divide toxic reports by a fixed-size time window, defined as 1,000 consecutive numeric identifiers of the toxic reports. We compute the proportion of toxic reports that come from allies to those from opponents in each time window and find that the number of reports by allies is consistently higher than

those by opponents, although it varies over time. This shows that frequent intra-group conflicts through chat are not artificial effects of the user interface.
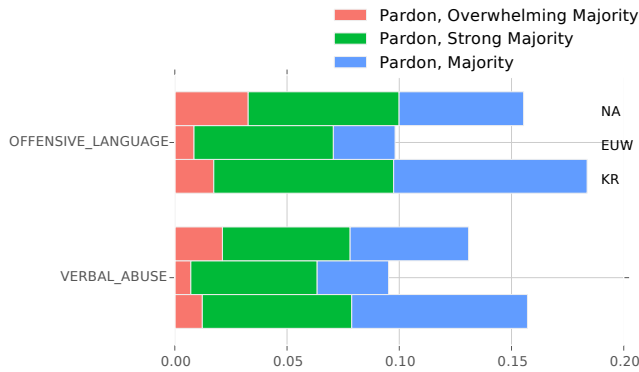

Figure 5. The percentage of cyberbullying reports that result in pardons for each region.

To test *[H3.1] Due to a more group-success oriented socio-political environment, cyberbullying offenses are less likely to be punished in Korea than in other regions*, we examine the rate of pardons for cyberbullying offenses in the different regions.

Figure 5 plots the percentage of cyberbullying reports (offensive language and verbal abuse) that result in pardons for each region. 17.1% of such reports are pardoned in KR, compared to 14.3% in NA and 9.7% in EUW. A Chi-Square test with Yates continuity correction reveals the effect of region on pardons in those categories is significant ($\chi^2(2, N = 3,108,172) = 24123.3, p < .0001$). Thus H3.1 is supported.

As we reviewed in previous section, a likely explanation for this is due to the *Wang-tta* concept in KR. Particularly invasive in gaming culture, *Wang-tta* probably leads to reviewers empathizing not with the cyberbullying victim, but rather the alleged toxic player who verbalized his displeasure with the victim's performance.

To test *[H3.2] Reports on toxic behavior that largely affects the result of the match are more often submitted in Korea than in other regions*, we compare the percentage of reports for intentional feeding or assisting enemy across regions since such toxic behavior directly contradicts the group-success goal.

We find support for H3.2 by looking at the mean and median of reports for assisting enemy or intentional feeding coming from teammates (1.482 and 1, 1.714 and 2, and 1.75 and 2 for NA, EUW, and KR, respectively). A Kruskal-Wallis test reveals the effect of region on reports coming from teammates is significant ($\chi^2(2) = 31175.43, p < .0001$). A post-hoc test using Mann-Whitney tests with Bonferroni correction confirms the significant differences between NA and EUW (p < .0001, r = .129), between NA and KR (p < .0001, r = .148), and between EUW and KR (p < .0001, r = .02).

We move on to testing *[H3.3] Reports on behavior that largely affects the result of the match are more likely to be punished in Korea than in other regions*. We plot the percentage of punish decisions for behavior that directly affects the
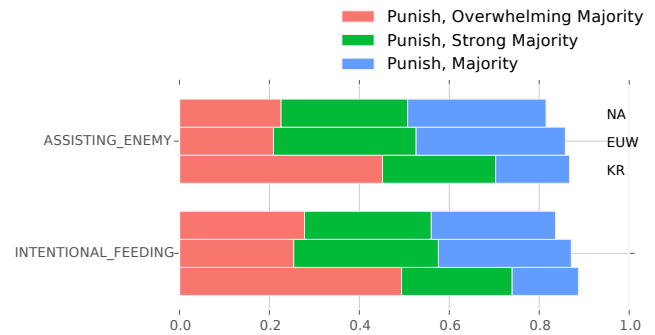

Figure 6. The percentage of reports for behavior that directly affects the result of the match for each region that results in a punishment.

result of the match for each region in Figure 6. While such offenses are heavily punished in each region (over 80%), we do find a difference. Korean reviewers are more likely to perceive assisting enemy and intentional feeding as severe toxic playing with much higher levels of agreement than other regions (for Punish, Overwhelming Majority, KR: 48%, NA: 27%, EUW: 24%), which is affirmative support for H3.3. A Chi-Square test with Yates continuity correction reveals the effect of region on the percentage of Punish, Overwhelming Majority in those categories is significant ($\chi^2(2, N = 1,955,297) = 83593.08, p < .0001$).
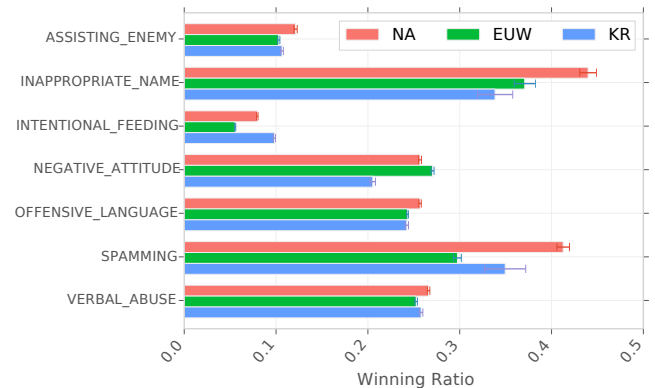
Team-cohesion and Performance

Figure 7. The winning ratio for each category of toxic behavior with 95% confidence interval.

To test *[H4.1] More reports come from matches where the accused was on the losing team*, we plot the winning ratio with 95% confidence interval for the different categories of toxic play reported in Figure 7. From the figure, it is immediately apparent that winning ratio is clearly below than 50%, even though LoL uses a match making system similar to Elo ranking [24] that attempts to match players in a manner where there is a 50% chance of winning for each team. In particular, we see that the winning ratio for intentional feeding and assisting the enemy are extremely low (under 15% for both). Therefore, more reports come from losing teams so that the observed winning ratio is less than 0.5. H4.1 is supported.

As we mentioned in the research question section, lower team cohesion leads to lower performance. Or, a poor performance might trigger toxic behavior like cyberbullying and also might

be a cause for the high number of reports involving a losing team. Cyberbullying offenses are explained by attribution theory in that when a toxic player recognizes a poor performance (e.g., even though the match is not decided, his team is losing) he looks for someone other than himself to place blame. Related to this, for example in the spamming category, players that were on the losing side of the match might attempt to attribute the loss to a another member of the team and attempt to punish him via the reporting system.
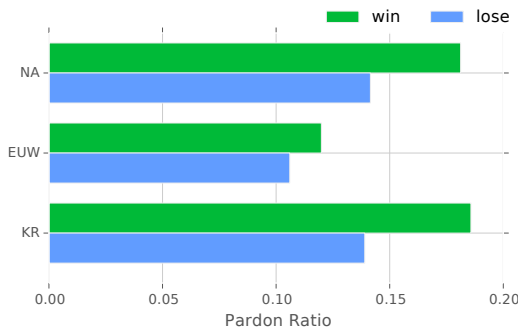

Figure 8. Pardoned ratio when losing vs. winning

To test *[H4.2] There are more cases pardoned when the accused was on the losing team than on the winning team*, we plot the pardoned ratio when the accused toxic player was on the winning or losing side. Figure 8 plots the breakdown of pardon and punish decisions as a function of whether the accused toxic player was on the winning or losing side. If people regard somewhat innocuous factors as significant reasons of defeat and report it as toxic behavior, the pardoned ratio for defeats will be higher than that for wins. As seen in Figure 8, by contrast, we find that the proportion of being pardoned by crowds when losing is less than that when winning. Thus, H4.2 is *not* supported, and in fact the opposite is the case.

DISCUSSION AND CONCLUSION

In this work we explored toxic playing and the reaction to it via crowdsourced decisions using a few million observed reports. Our large-scale dataset enables an opportunity to explore several compelling theories from sociology and psychology that discuss the "whys" of toxic behavior.

We first showed that there is relatively low participation in reporting toxic behavior and reconfirmed the impact of anonymity in CMC and cyberspace. This finding underlines the difficulties in relying solely on voluntary player reports. For example, in Facebook there is a button to report a comment that violates cyber-etiquette. The low degree of participation for social control we found raises a fundamental question about the design considerations of such report-based systems: if relatively few "victims" voluntarily report such behavior, how effective can it truly be? Interestingly, our finding that an explicit request to report toxic behavior significantly increases the likelihood of reporting indicates that actively encouraging reporting should be considered in the design of systems to address toxic behavior.

Next, we examined the vague nature of toxic behavior. This issue is repeatedly observed in online games [34]. With over 10% of cases being pardoned in the Tribunal, it is clear that crowdsourcing using experienced players is useful in protecting innocent victims who are wrongly reported by other players due to poor gaming skills or aggressive (but not toxic) linguistic behavior. The Tribunal could further improve the quality and efficiency of crowdsourced workers via several proposed mechanisms for quality control in crowdsourced systems and augmentation with machine-learning solutions [11].

We then moved on to how group setting might influence reporting behavior. We quantitatively show that in-group favoritism and out-group hostility increase or decrease the willingness to report. This is rooted in competition among players. Since competition is a common game design element for enjoyment [52], our findings are applicable to most online games. Even though a given game might not have a form of team competitions, sense of belonging is frequently derived from a small group of players, called guilds or parties in MMORPGs [22]. In a broad sense, homophily can appear not only in an explicit group but also in an implicit group [37]. The current work portends possible bias of observed reports in such settings.

Most online games allow interactions between players. In this environment toxic playing is a serious issue that degrades user experience. Our work offers understanding of toxic playing and its victims based on LoL data, but much of the mechanics involved are typical game elements not unique to LoL. We also believe that with the growing trend of gamification (e.g., as applied to citizen science) our findings have broader application than traditional entertainment gaming. In fact, some gamified citizen science projects have already experience low level toxicity [7]. As designers and scientists import more and more gaming elements into their systems, they will most assuredly be accompanied by less desirable elements of gaming culture, such as toxic behavior.

We also see similarities between LoL and online communities. Online communities, e.g., Reddit, allow anonymous user identities. They use unique user names which are not linked to real identity. Although there are differences between forms of toxic playing and cyberbullying in online communities, the disconnect between real and virtual world is a common root that both cyberbullying and toxic playing stem from.

Beyond gaming, our work deepens understanding of team conflicts in general. This has great potential because teams are an essential building block of modern organizations. Also, in the current globally distributed workspace, collaboration through an electronic channel is pervasive. This often results in a goal-oriented group that lacks *social* connections. Such settings can accelerate the tendency to blame others, further escalated by the individuals' unfamiliarity with remote partners [19]. Competitive games in general, and the LoL Tribunal in particular, are thus a valuable asset to capture group conflicts in the virtual space and test effective treatments for them. We believe that solutions for toxic play in team com-

petition online games could have huge impact for real-world scenarios, and not just virtual spaces.

For the overall CHI community, while ethnography is traditionally considered as one of the best methods to understand people, we show that studying human behavior with big data and testable hypotheses works well. We hope that this helps accelerate the discovery of big data's value by the CHI community.

Caveats and Limitations
Although our findings confirm several theories on cyberbullying and toxic behavior, we note that there are limitations and caveats that must be considered. First, although there is reason to believe that gamers generally behave the same in-game as they do in the real-world [12], the fact remains that our dataset is drawn from a game. At minimum, this means applying our results to other domains must be done carefully [30]. Further, our findings are from LoL in particular, and there are domain specific concerns that might not be present in other games even within the same genre. For example, SMITE is currently the 3rd most popular MOBA in the market, but completely lacks an all-talk chat mode. Additionally, SMITE and Dota 2 (the 2nd most popular MOBA) both have integrated voice communication, which might intensify or soften cyberbullying with feelings of social presence.

Next, the social network structure among players, "friend" relationships, is not considered in this work due to lack of data. We thus do not incorporate social relationship among specific players in building our hypotheses. As recent studies reveal that playing together with friends influences performance and toxicity [36, 45], we believe that more detailed data including players' social networks would enable testing of more sophisticated hypotheses.

Finally, although our dataset is large-scale and quite rich, it is anonymized and thus prevents us from knowing how toxic players behaved before and after the matches were aggregated into their case and a decision being made. We also lack knowledge of how other players in the game typically behave. Thus, while we were able to provide significant support for several theories, there are questions for which answers continues to elude us.

REFERENCES
1. IWillDominate Tribunal permaban & eSports competition ruling. http://goo.gl/2qHH4I.

2. Player tally for League of Legends surges. http://goo.gl/YxjSjb.

3. Reporting a player. https://support.riotgames.com/hc/en-us/articles/201752884-Reporting-a-Player.

4. Riot Games' League of Legends officially becomes most played pc game in the world. http://goo.gl/iqrru6.

5. Riot Games releases awesome League of Legends infographic. http://goo.gl/bDKLnO.

6. Suicide of Megan Meier. http://goo.gl/MHgLJG.

7. To anyone who seen the recent popular post about a game called Foldit. This is why you should avoid it. http://redd.it/2n97b1.

8. Aggression, R., and Roles, P. M. Cyberbullying in Japan. *Cyberbullying in the Global Playground: Research from International Perspectives* (2011), 183.

9. Barlińska, J., et al. Cyberbullying among adolescent bystanders: role of the communication medium, form of violence, and empathy. *Journal of Community & Applied Social Psychology 23*, 1 (2013), 37–51.

10. Blackburn, J., et al. Branded with a scarlet "C": cheaters in a gaming social network. In *WWW* (2012).

11. Blackburn, J., and Kwak, H. STFU NOOB! predicting crowdsourced decisions on toxic behavior in online games. In *WWW* (2014).

12. Boellstorff, T. *Coming of age in Second Life: An anthropologist explores the virtually human*. Princeton University Press, 2008.

13. Carron, A. V. Cohesion and performance in sport. *Journal of Sport & Exercise Psychology 24* (2002), 168–188.

14. Chee, F. The games we play online and offline: Making Wang-tta in Korea. *Popular Communication 4*, 3 (2006), 225–239.

15. Chen, V. H.-H., Duh, H. B.-L., and Ng, C. W. Players who play to make others cry: The influence of anonymity and immersion. In *ACE* (2009).

16. Chesney, T., et al. Griefing in virtual worlds: causes, casualties and coping strategies. *Information Systems Journal 19*, 6 (2009), 525–548.

17. Christopherson, K. M. The positive and negative implications of anonymity in Internet social interactions:"on the Internet, nobody knows you're a dog". *Computers in Human Behavior 23*, 6 (2007), 3038–3056.

18. Cox, T. H., Lobel, S. A., and McLeod, P. L. Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of management journal 34*, 4 (1991), 827–847.

19. Cramton, C. D. The mutual knowledge problem and its consequences for dispersed collaboration. *Organization science 12*, 3 (2001), 346–371.

20. Denant-Boemont, L., Masclet, D., and Noussair, C. N. Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic theory 33*, 1 (2007), 145–167.

21. Diener, E. Deindividuation, self-awareness, and disinhibition. *Journal of Personality and Social Psychology 37*, 7 (1979), 1160.

22. Ducheneaut, N., et al. The life and death of online gaming communities: A look at guilds in World of Warcraft. In *CHI* (2007).

23. Einhorn, H. J., and Hogarth, R. M. Judging probable cause. *Psychological Bulletin 99*, 1 (1986), 3.

24. Elo, A. E. *The rating of chessplayers, past and present*, vol. 3. Batsford London, 1978.

25. Felps, W., Mitchell, T. R., and Byington, E. How, when, and why bad apples spoil the barrel: Negative group members and dysfunctional groups. *Research in organizational behavior 27* (2006), 175–222.

26. Foo, C. Y., and Koivisto, E. M. I. Defining grief play in MMORPGs: player and developer perceptions. In *ACE* (2004).

27. Hofstede, G. H. *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage, 2001.

28. Huang, Y., et al. Functional or social?: exploring teams in online games. In *CSCW* (2013), 399–408.

29. Kavussanu, M., and Ntoumanis, N. Participation in sport and moral functioning: Does ego orientation mediate their relationship? *Journal of Sport and Exercise Psychology 25* (2003), 501–518.

30. Keegan, B., et al. Sic transit gloria mundi virtuali?: promise and peril in the computational social science of clandestine organizing. In *WebSci* (2011).

31. Kirman, B., et al. Exploring mischief and mayhem in social computing or: how we learned to stop worrying and love the trolls. In *CHI EA* (2012).

32. Kwak, H., and Blackburn, J. Linguistic analysis of toxic behavior in an online video game. In *EGG workshop* (2014).

33. Latane, B., and Darley, J. M. *The unresponsive bystander: Why doesn't he help?* Appleton-Century Crofts New York, 1970.

34. Lin, H., and Sun, C.-T. The "White-Eyed" player culture: Grief play and construction of deviance in MMORPGs. In *DiGRA Conference* (2005).

35. Marsden, P. V., and Campbell, K. E. Measuring tie strength. *Social forces 63*, 2 (1984), 482–501.

36. Mason, W., and Clauset, A. Friends FTW! friendship and competition in Halo: Reach. In *CSCW* (2013).

37. McPherson, M., Smith-Lovin, L., and Cook, J. M. Birds of a feather: Homophily in social networks. *Annual review of sociology* (2001), 415–444.

38. Mullen, B., and Copper, C. The relation between group cohesiveness and performance: An integration. *Psychological bulletin 115*, 2 (1994), 210.

39. Naito, T., and Gielen, U. Bullying and Ijime in Japanese schools. In *Violence in Schools*. Springer, 2005.

40. Naquin, C. E., and Tynan, R. O. The team halo effect: Why teams are not blamed for their failures. *Journal of Applied Psychology 88*, 2 (2003), 332.

41. Ommundsen, Y., Roberts, G., Lemyre, P., and Treasure, D. Perceived motivational climate in male youth soccer: Relations to social–moral functioning, sportspersonship and team norm perceptions. *Psychology of Sport and Exercise 4*, 4 (2003), 397–413.

42. Paulhus, D. L. *Measurement and control of response bias.* Academic Press, 1991.

43. Reicher, S. D., Spears, R., and Postmes, T. A social identity model of deindividuation phenomena. *European Review of Social Psychology 6*, 1 (1995), 161–198.

44. Roese, N. J. Counterfactual thinking. *Psychological bulletin 121*, 1 (1997), 133.

45. Shores, K. B., He, Y., Swanenburg, K. L., Kraut, R., and Riedl, J. The identification of deviance and its impact on retention in a multiplayer game. In *CSCW* (2014).

46. Smith, P. K., et al. Cyberbullying: Its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry 49*, 4 (2008), 376–385.

47. Szell, M., Lambiotte, R., and Thurner, S. Multirelational organization of large-scale social networks in an online world. *PNAS 107*, 31 (2010), 13636–13641.

48. Thompson, P. What's fueling the flames in cyberspace? a social influence model. *Communication and cyberspace: Social interaction in an electronic environment* (1996), 293–311.

49. Tönnies, F. *Gemeinschaft und gesellschaft*. Springer, 2012.

50. Triandis, H. C., et al. Individualism and collectivism: Cross-cultural perspectives on self-ingroup relationships. *Journal of personality and Social Psychology 54*, 2 (1988), 323.

51. Turner, J. C. Social comparison and social identity: Some prospects for intergroup behaviour. *European journal of social psychology 5*, 1 (1975), 1–34.

52. Vorderer, P., Hartmann, T., and Klimmt, C. Explaining the enjoyment of playing video games: The role of competition. In *ACE* (2003).

53. Warner, D. E., and Ratier, M. Social context in massively-multiplayer online games (MMOGs): Ethical questions in shared space. *International Review of Information Ethics 4*, 7 (2005).

54. Weibel, D., et al. Playing online games against computer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior 24*, 5 (Sept. 2008), 2274–2291.

55. Weiner, B. A cognitive (attribution)-emotion-action model of motivated behavior: An analysis of judgments of help-giving. *Journal of Personality and Social Psychology 39*, 2 (1980), 186.